Metabarcoding for the parallel identification of several hundred predators and their preys: application to bat species diet analysis

MAXIME GALAN✉[1], JEAN-BAPTISTE PONS[2], ORIANNE TOURNAYRE[1], ERIC PIERRE[1], MAXIME LEUCHTMANN[3], DOMINIQUE PONTIER[2,4,*] and NATHALIE CHARBONNEL[1*]

Full postal addresses

[1]CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France

[2]Univ Lyon, LabEx ECOFECT Ecoevolutionary Dynamics of Infectious Diseases, Villeurbanne F-69365 Lyon, France

[3] Nature Environnement, 2 Avenue Saint-Pierre, 17700 Surgères, France

[4]Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR5558, F-69622 Villeurbanne, France

* equal author contributions

Keywords (4-6): Chiroptera, arthropods, environmental DNA (eDNA), high-throughput sequencing, false positives, predator–prey interactions

Name, address, fax number and email of corresponding author

✉ Maxime Galan, INRA, UMR CBGP (INRA/IRD/CIRAD/Montpellier Supagro/Univ. Montpellier), 755 avenue du Campus Agropolis, CS 300 16, F-34988 Montferrier-sur-Lez cedex, France. Fax: (+00 33) 04 99 62 33 05; E-mail: maxime.galan@inra.fr

Running title : Metabarcoding of predators and their preys

Abstract

Assessing diet variability between species, populations and individuals is of main importance to better understand the biology of bats and design conservation strategies. Although the advent of metabarcoding has facilitated such analyses, this approach does not come without challenges. Biases may occur throughout the whole experiment, from fieldwork to biostatistics, resulting in the detection of false negatives, false positives or low taxonomic resolution. We detail a rigorous metabarcoding approach based on a two-step PCR protocol enabling the 'all at once' taxonomic identification of bats and their arthropod preys for several hundreds of samples. Our study includes faecal pellets collected in France from 357 bats representing 16 species, as well as insect mock communities that mimic bat meals of known composition, negative and positive controls. All samples were analysed in triplicate. We compare the efficiency of DNA extraction methods and we evaluate the effectiveness of our protocol using the rate of bat identification success, taxonomic resolution, sensitivity, and amplification biases. Our strategy involves twice fewer steps than usually required in the other metabarcoding studies and reduces the probability to misassign preys to wrong samples. Controls and replicates enable to filter the data and limit the risk of false positives, hence guaranteeing high confidence results for both prey occurrence and bat species identification. Our results illustrate the power of this approach to assess diet richness and variability within and between colonies. We therefore provide a rapid, resolutive and cost-effective screening tool for addressing evolutionary ecological issues or developing 'chirosurveillance' and conservation strategies.

Introduction

DNA metabarcoding, the 'high-throughput multi-species (or higher-level taxon) identification using the total and typically degraded DNA extracted from an environmental sample' (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012), has revolutionized our approaches of biodiversity assessment this last decade. The method is based on the high-throughput sequencing (HTS) of DNA barcode regions (i.e. small fragments of DNA that exhibit low intra-species and high inter-species variability), which are amplified using universal PCR primers (Hebert, Cywinska, Ball, & deWaard, 2003). Nowadays recent HTS technologies (e.g. Illumina) generate millions of sequences concurrently. Metabarcoding therefore enables to characterize quickly and in a single reaction a very large number of species present in an environmental sample, and also to analyse simultaneously several hundreds of samples by using tags / index and multiplexing protocols (Binladen et al., 2007). Metabarcoding has proved to be useful for a wide variety of applications, including the analysis of aquatic and terrestrial microbiome ecosystems (e.g. Zinger et al., 2011), or the taxonomic identification of animal and vegetal species in complex samples including food, faeces, permafrost, soil or water (Bohmann et al., 2014; Taberlet et al., 2012).

Dietary analyses have been facilitated by the advent of metabarcoding and its application to the analysis of faeces or stomach contents (Pompanon et al., 2012). Compared to traditional morphological analyses of remaining hard parts, a large sampling can be processed quickly. Results are more sensitive, allowing the identification of a larger array of specimens (e.g. juvenile life stages) and providing a greater taxonomic resolution (e.g. cryptic species might be detected, Hebert, Penton, Burns, Janzen, & Hallwachs, 2004). In addition, traditional molecular approaches (real time PCRs, Sanger sequencing) require several assays or reactions to discriminate the

different taxa present in the same sample, while a single run of metabarcoding provides the identification of a broad range of taxa with no *a priori* (Tillmar, Dell'Amico, Welander, & Holmlund 2013). Research on insectivorous bat dietary analyses have been pioneering in benefiting from these advantages of DNA metabarcoding (Bohmann et al., 2011). Indeed, direct observations of prey consumption are made difficult as these species fly and are nocturnal. Moreover, because many bat species are vulnerable or endangered around the world, catching might be difficult, even forbidden during hibernation, and invasive methods cannot be applied. Morphologic examinations of faeces and guano have therefore initially provided important knowledge on bat diets (Hope et al., 2014; Lam et al., 2013), but these methods have major limits (time consuming, taxonomic expertise required, low resolution and ascertainment biases due to the reject of insect hard parts, see refs in Iwanowicz et al., 2016). In particular, identifying preys at the species level is not possible based on morphological analyses of faecal samples, although this taxonomic information may be of main importance for conservation, agronomical or Public Health applications. Indeed, the foraging activity of insectivorous bats can strongly reduce the abundance of crop pests (Maine & Boyles, 2015; McCracken et al., 2012), some of them vectors of infectious diseases (Gonsalves, Law, Webb, & Monamy, 2013). Emphasizing these ecosystem services provided by bats in natural or agricultural ecosystems is of main importance to define conservation priorities and practices (Clare, 2014), and it should be based on a accurate identification of preys, as phylogenetically close arthropod species might not all be enemies.

Besides, molecular approaches do not come without challenges. Obtaining reliable and reproducible results from metabarcoding is not straightforward as several biases may occur throughout the whole experiment, resulting in the detection of false negatives, false positives or to low taxonomic resolution (Ficetola, Taberlet, & Coissac,

2016). These biases take place from fieldwork to biostatistics (see for a review in an epidemiological context, Galan et al., 2016). Contaminations occurring during sampling or in the laboratory (Champlot et al., 2010 ; Goldberg et al., 2016) may be further amplified and sequenced due to the high sensitivity of the PCR and to the high depth of sequencing provided by the HTS, leading to further misinterpretations (see examples in Ficetola et al., 2016). The choice of the DNA extraction method, the barcode region and primers may also influence the issue of metabarcoding studies. Low efficiencies of sample disruption, high losses of genetic material or the presence of PCR inhibitors may lead to false negative results (Deiner, Walser, Machler, & Altermatt, 2015). Incorrect design of primers and unsuitable barcode region may prevent or bias the amplification of the taxonomic taxa studied, or may result in an identification at low resolution levels (Hajibabaei et al., 2006). In addition to these well-known precautions required for metabarcoding studies, other considerations need to be made. First, mutliplexing samples within HTS runs results in mis-assignments of reads after the bioinformatic demultiplexing (Kircher, Sawyer, & Meyer, 2012). The detection of one or a few reads assigned to a given taxon in a sample therefore does not necessarily mean this taxon is actually present in that sample (Galan et al., 2016). These errors may originate from: (i) contamination of tags/index, (ii) production of between samples chimera due to jumping PCR (Schnell, Bohmann, & Gilbert, 2015) when libraries require the bulk amplification of tagged samples, or (iii) the presence of mixed cluster of sequences (i.e. polyclonal clusters) on the Illumina flowcell surface. They may have dramatic consequences on the proportion of false positives (e.g. up to 28.2% of mis-assigned unique sequences reported in Esling, Lejzerowicz, & Pawlowski, 2015). Unfortunately, read mis-assignments due to polyclonal clusters during Illumina sequencing are difficult to avoid and concern 0.2 to 0.6% of the reads generated (Galan et al., 2016 ; Kircher et

al., 2012 ; Wright & Vetsigian, 2016). It is therefore of main importance to filter the occurrence data obtained through metabarcoding experiments, using both controls and replicates (Ficetola et al., 2016; Galan et al., 2016; Robasky, Lewis, & Church, 2014). The second set of parameters still scarcely considered during metabarcoding experiments includes the sensitivity and taxonomic resolution of the protocol designed. They can be assessed empirically by analysing mock communities (MC), *i.e.* pools of DNA belonging to different species, hence simulating a predator meal of known composition (Pinol, Mir, Gomez-Polo, & Agusti, 2015).

In this study, we propose a rigourous metabarcoding approach based on a two-step PCR protocol and bioinformatic analyses enabling the '*all at once*' identification, potentially at the species level, of bats and their arthropod preys for several hundreds of samples. We use faecal pellets from 357 bats representing 16 species collected in Western France in 2015. Our aims are threefold. First, we compare the efficiency of DNA extraction and purification methods among six available commercial kits. Second, we design a scrupulous experimental protocol that includes negative and positive controls as well as systematic technical replicates. They enable to filter occurrence results, in particular with regard to false positives (Ficetola et al., 2016; Galan et al., 2016). Then, we evaluate the effectiveness of this protocol using a set of criteria including the rate of identification success, taxonomic resolution, sensitivity, and amplification biases. To this end, we analyse arthropod mock communities to mimic bat meals of known composition, and we validate predator identifications by comparing molecular results with the individual morphological identifications performed during fieldwork by experts. Third, we apply this DNA metabarcoding protocol to identify the consumed preys of the 357 faecal bat samples collected on the field, and we examine our results with regard to morphological-based dietary analyses of bats previously published in the

literature.

Materials and methods

Study sites and sample collection

Bats were captured from summer roost sites (parturition colonies and swarming sites) between June and September 2015 in 18 sites located in Western France (Poitou-Charentes). For each site, harp traps were placed one night, at the opening of cave or building before sunset. Each captured bat was placed in a cotton holding bag until it was weighted, sexed and measured. Species identification was determined based on morphological criteria. Bats were then released. Faecal pellets were collected from holding bag, and stored in 2mL microtubes at room temperature until DNA was extracted several months later. Authorization for bat capture was provided by the Ministry of Ecology, Environment, and Sustainable development over the period 2015-2020. All methods were approved by the MNHN and the SFEPM.

DNA extraction from faecal samples

We analysed faecal pellets from 357 bats corresponding to 16 species of Chiroptera. Details are provided in Table S1. One pellet per individual was extracted. We randomised the 357 faecal samples between six silica-membrane DNA extraction kits to compare their efficiency: EZ-10 96 DNA Kit, Animal Samples (BioBasic; n = 113), QIAamp Fast DNA Stool Mini Kit (Qiagen; n = 47), DNeasy mericon Food Kit (Qiagen; n = 47), ZR Fecal DNA MiniPrep (Zymo; n = 46), NucleoSpin 8 Plant II (Macherey-Nagel; n = 95) and NucleoSpin Soil (Macherey-Nagel; n = 9). The EZ-10 96 DNA and NucleoSpin 8 Plant II kits provide high-throughput DNA isolation (up to 192 samples in parallel) thanks to a 96-well format unlike the other kits using tube format. For all kits, we

followed manufacturer's recommendations except for the NucleoSpin 8 Plant II as we used the protocol improved by E. Petit (pers. comm.).

We compared DNA extraction kits' efficiency using four criteria: i) the mean number of reads per PCR obtained after sequencing; ii) the proportion of failure during sequencing (PCR with less than 500 reads); iii) the success rate of host sequencing (presence of chiropter reads from the same unique sequence/haplotype, found repeatedly between the PCR triplicates); and iv) the success rate of prey sequencing (presence of unique sequences corresponding to arthropods, found repeatedly between the PCR triplicates).

Mock community preparation

To better evaluate the sensitivity of our metabarcoding approach, we created two artificial communities of arthropods that mimic insectivorous bat diets. The first mock community (MC_1) was composed of 12 species collected in France, including four Coleoptera, two Dermaptera, two Hemiptera, one Lepidoptera, two Neuroptera and one Orthoptera (Table 1). The second mock community (MC_2) included seven species collected worldwide, with two Coleoptera, one Diptera, one Hemiptera, two Lepidoptera and one Orthoptera (Table 1).

Arthropod DNA was extracted individually using the DNeasy Blood & Tissue kit (Qiagen). Sanger sequences were available for the COI gene of each individual (see Suppl. Mat. Fasta S1). DNA extractions were normalized to 5ng/µL using Qubit fluorimeter quantification (Invitrogen). Each DNA was amplified and sequenced independently on one hand, and mixed at equal concentration with all DNA extracts corresponding to a given mock community before amplification and sequencing on the other hand (see detailed in Table S1).

Laboratory precautions and controls

Throughout the experiment, we strictly applied the laboratory protocols to prevent contamination by alien DNA and PCR products. All pre-PCR laboratory manipulations were conducted with filter tips under a sterile hood in a DNA-free room, i.e., a room dedicated to the preparation of PCR mix and equipped with hoods that are kept free of DNA by UV irradiation and bleach treatment. The putative presence of contamination was checked at this stage and along the whole laboratory procedure using different negative and positive controls. A large number of research has highlighted several biases occurring at different steps of amplicon HTS (for a detailed list, see Galan et al., 2016). These biases can be estimated directly from the data by including several controls together with samples in the experiment. They are explained below.

Negative Controls for DNA extraction (NC_{ext}) are DNA extractions performed without any sample tissue (blank). For each DNA extraction session, one NC_{ext} is processed with the other samples. The numbers of sequences observed in NC_{ext} are used to detect and filter the cross-contaminations during DNA extractions and the next steps.

Negative Controls for PCR (NC_{PCR}) are PCR reactions performed without any DNA extract (blank). They are processed together with the other samples. One NC_{PCR} is performed per 96-well PCR microplate. The numbers of sequences observed in NC_{PCR} are used to detect and filter the cross-contaminations during PCR preparation and the next steps.

Negative Controls for indexing (NC_{index}) are combinations of multiplexing indexes that are not included to identify samples in the HTS run, but that are searched for during the bioinformatic demultiplexing (Esling, Lejzerowicz, & Pawlowski, 2015). They correspond to empty PCR wells (*i.e.* neither reagent nor index pair). The numbers of sequences recovered for these particular index pairs are used to detect and filter the

cross-contaminations between indexed PCR primers during primer handling or indexing PCR preparation, or to identify errors in the Illumina sample sheet.

Positive Controls for PCR (PC_{PCR}) are PCR reactions with DNA of known taxa that are processed together with the other samples. The sequences obtained for these controls are used to verify the success of the experiment and the taxonomic identification. In this experiment, the 19 arthropods of the mock communities (MC_1 & MC_2) performed in individual PCRs and in pools are used as PC_{PCR} (see "Mock community preparation" section).

Positive Controls for Indexing (PC_{alien}) are particular PC_{PCR} that correspond to PCR reactions with DNA of taxa that are known to be absent in the samples. They are handled separately from the samples to avoid cross-contaminations with these latter during the wet lab procedures (DNA extractions and PCRs). Sequences from PC_{alien} found in the other samples are used to calculate the rate of read mis-assignments due to false index-pairing generated predominantly by polyclonal clusters during Illumina sequencing (see Kircher et al. 2012 for details concerning this phenomenon). In this experiment, nine PCRs of DNA extracted from beluga whale (*Delphinapterus leucas*) tissue are used to estimate the rate of sample misidentification. This rate is used to filter out the data. We strongly recommend to use taxa that are phylogenetically distant from the taxa of interest (here Chiroptera and Arthropoda), in order to avoid potential confusion between sequences from alien controls and sequences from the samples. Details of the positive and negative controls included in this experiment are provided in Table S1.

PCR and library construction

We performed a two-step PCR strategy (see Illumina Application Note Part 15044223) combined with the dual-index paired-end sequencing approach described in Kozich,

Westcott, Baxter, Highlander, and Schloss (2013; Fig. 1). The 32 index i5 and 48 index i7 allow to multiplex up to 1536 PCR products in the same MiSeq run.

During the first PCR (PCR_1), we used the 133 bp minibarcode of cytochrome oxydase I (COI) described in Gillet et al. (2015). The primer pair corresponded to modified versions of forward primer LepF1 (Hebert et al., 2004) and reverse primer EPT-long-univR (Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011) as its efficiency to identify arthropods from France has already been proven (Gillet et al. 2015). We added the partial overhang Illumina sequencing primers in 5' end and an heterogeneity spacer of each target-specific primer (Figure 1): LepF1-MiSeq 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG(none/C/GC/TGC/CTGC/TCCGG) ATTCHACDAAYCAYAARGAYATYGG-3' and EPT-long-univR-MiSeq 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG(none/G/CT/TGG/CGCG/GCGTG) ACTATAAAARAAAYTATDAYAAADGCRTG-3'. The alternative bases between the partial adaptors and the target-specific primers correspond to a 0 to 5 bp "heterogeneity spacer" designed to mitigate the issues caused by low sequence diversity in Illumina amplicon sequencing (Fadrosh et al., 2014). These five versions of each forward and reverse primer were mixed together before PCR_1. During the first cycles of the Illumina sequencing, they created an artificial diversity of the four nucleobases to improve the detection of the sequencing clusters at the flowcell surface, what consequently increased the quality of the reads. This PCR_1 was performed in 11 µL reaction volume using 5 µL of 2× Qiagen Multiplex Kit Master Mix (Qiagen), 0.5 µM of each mix of forward and reverse primers, and 2 µL of DNA extract. The PCR conditions consisted in an initial denaturation step at 95°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 45°C for 45 s, and extension at 72°C for 30 s, followed by a final extension step at 72°C for 10 min.

During the second PCR (PCR_2), we used the PCR_1 as DNA template. This PCR_2 consists of a limited-cycle amplification step to add multiplexing indices i5 and i7 and Illumina sequencing adapters P5 and P7 at both ends of each DNA fragment. The indexed primers P5 (5'-AATGATACGGCGACCACCGAGATCTACAC(8-base i5 index)TCGTCGGCAGCGTC-3') and P7 (5'-CAAGCAGAAGACGGCATACGAGAT(8-base i7 index)GTCTCGTGGGCTCGG-3') were synthetized with the different 8-base index sequences (Kozich, Westcott, Baxter, Highlander, & Schloss, 2013). PCR_2 was carried out in a 11 µL reaction volume using 5 µL of Qiagen Multiplex Kit Master Mix (Qiagen) and 0.7 µM of each indexed primer. To facilitate the distribution of indexed primers, we prepared stock solutions of P5 and P7 primer (3.5µM) in 8-well and 12-well tube strips respectively and we dispensed 2 µL of each primer in column or row of the each 96-well PCR microplate using a multichannel micropipette. Then, in a post-PCR room, 2 µL of PCR_1 product was added to each well. The PCR_2 started by an initial denaturation step of 95°C for 15 min, followed by 8 cycles of denaturation at 94°C for 40 s, annealing at 55°C for 45 s and extension at 72°C for 60 s followed by a final extension step at 72°C for 10 min.

PCR_2 products (3 µL) were verified by electrophoresis in a 1.5% agarose gel. One PCR blank (NC_{PCR}) and one negative control for indexing (NC_{index}) were systematically added to each of the 15 PCR microplate. Each DNA extraction was amplified and indexed in three independent PCR reactions. These PCR triplicates were used as technical replicates to confirm the presence of a taxa in a sample and further remove the false positive results (Robasky et al., 2014). A MiSeq (Illumina) run was conducted, including PCR_2 products from bat faecal samples (number of PCRs n = 357 x 3), the positive (n = 24 x 3 including arthropods used in the mock communities) and negative (n = 58) controls (see details in Table S1).

MiSeq sequencing

The MiSeq platform was chosen because it generates lower error rates than other HTS platforms (D'Amore et al., 2016). For this study, the number of PCR products multiplexed was 1168 (Table S1). PCR products were pooled by volume for each 96-well PCR microplate. Mixes were checked by electrophoresis in 1.5% agarose gel before generating a 'super-pool' including all PCR products. We subjected 60 µL of the super-pool to size selection for the full-length amplicon (expected size: 312 bp including primers, indexes and adaptors) by excision on a low-melting agarose gel (1.25%). It enabled to discard non-specific PCR products and primer dimers. The PCR Clean-up Gel Extraction kit (Macherey-Nagel) was used to purify the excised bands. The super-pool of amplicon library was quantified using the KAPA library quantification kit (KAPA Biosystems) before loading 8 pM and 10% of PhiX control on a MiSeq flow cell (expected cluster density: 700-800 K/mm2) with a 500-cycle Reagent Kit v2 (Illumina). We performed a run of 2 x 200 bp paired-end sequencing, which yielded high-quality sequencing through the reading of each nucleotide of the COI minibarcode fragments twice after the assembly of reads 1 and reads 2. Information about PCR products and fastq file names are provided in Table S1.

Sequence analyses and data filtering

We used MOTHUR program v1.34 (Schloss et al., 2009) to create a table of abundance for each unique sequence (i.e. dereplicated sequences) and each PCR product. Hereafter, a unique sequence will correspond to a cluster of several identical reads. Briefly, MOTHUR enabled to i) contig the paired-end read 1 and read 2; ii) remove the reads with low quality of assembling (> 200 bp); iii) dereplicate the redundant reads in unique sequence; iv) align the unique sequences on a COI minibarcode reference alignment; v)

remove PCR primers; vi) remove misaligned unique sequences; vii) cluster the unique sequences that are within one substitution of each other; viii) remove the singleton and rare unique sequences (*cut-off* = 8 reads) and ix) remove chimeric unique sequences using UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011).

Because multiple biases may occur during sample preparation and data processing (see details in Galan et al., 2016), it is impossible to distinguish true and false positives for particular hosts/preys directly from the table of abundance provided by MOTHUR. We here propose a conservative approach enabling to filter the data for putative false positive results. We used the multiple controls introduced during the process to estimate the potential different biases and define read number thresholds above which the PCR product may be considered as positive for a given sequence. Following Galan et al. (2016), two different thresholds were set for each unique sequence, and a cross-validation using the PCR triplicates was applied to confirm the positivity of each sample for each unique sequence.

First, T_{CC} threshold was used to filter cross-contaminations during the laboratory procedure. For each unique sequence, we used the maximum number of reads observed in the different negative controls (NC) as threshold. PCR products with fewer than this number of reads for this particular unique sequence were considered to be negative. Second, T_{FA} threshold was applied to filter the false-assignments of reads to a PCR product due to the generation of mixed clusters during the sequencing (Kircher et al., 2012). This phenomenon was estimated in our experiment using "alien" positive controls (PC_{alien}) sequenced in parallel with the bat faecal samples. As the PC_{alien} were handled separately from bat samples before the sequencing, the presence of reads from beluga whale in a bat sample indicated a sequence assignment error due to the Illumina sequencing. We determined the maximal number of reads of beluga whale assigned to a

bat PCR product. We then calculated the false-assignment rate (R_{FA}) for this PCR product by dividing this number of reads by the total number of reads from beluga in the sequencing run. Moreover, the number of reads for a specific unique sequence mis-assigned to a PCR product should increase with the total number of reads of this unique sequence in the MiSeq run. We therefore defined T_{FA} threshold as the total number of reads in the run for a unique sequence multiplied by R_{FA}. PCR products with fewer than T_{FA} for a particular unique sequence were considered to be negative. We then discarded positive results associated with numbers of reads below the thresholds T_{CC} and T_{FA}. Lastly, we discarded not-triplicated positive results.

Taxonomic assignment

We used BOLD Identification System {http://www.boldsystems.org/, \Ratnasingham, 2007 #2413} and species level barcode records (2,695,529 sequences/175,014 species in January 2017) to provide a taxonomic identification of each unique sequence passing our filtering processes. We provided a 5-class criteria describing the confidence level of the sequence assignments, modified from Razgour et al. (2011). They were applied to hits with similarity higher than 97% (see details in Table S2). For multi-taxonomic affiliation, we kept the common parent of all possible taxa. For sequences exhibiting similarity results lower than 97% in BOLD, we performed a BLAST in GENBANK to improve the taxonomic identification. Finally, results with similarity lower than 97% in BOLD and GENBANK were assigned to the phylum level using the closest taxa, or were considered as unclassified taxa when no match was found in the databases.

Statistical analyses

We emphasized the potential of this metabarcoding approach for addressing ecological issue by analysing the variability of bat diets among species and sites. We first applied a nonmetric multidimensional scaling (NMDS) using the Jaccard distance matrix built on the arthropod presence/absence (order level) of bat samples. Because bat species sampling was not balanced between sites, we considered independently two sites with large sample size, i.e. 'Jonzac' and 'Saint-Bonnet-sur-Gironde'. We presented two-dimensional ordination plots to visualize diet similarity/dissimilarity between individuals and bat species. Second, we applied a generalized linear model (GLM) with quasi-Poisson error structure (log link function) to evaluate the influence of bat species and sites on diet richness. Post-hoc Tukey tests were performed using the MULTCOMP libray in R.

Results

Sequencing results & data filtering

The MiSeq sequencing of 1162 PCR products including bat samples, positive and negative controls analysed in PCR triplicates generated a total paired-end read output of 6,918,534 reads of the COI minibarcode. MOTHUR program removed seven negative controls because they produced less than 8 reads, 633,788 (9.2%) of paired-end reads because they were misassembled, 312,336 (4.5%) of reads because they were misaligned, 125,606 (1.8%) of reads because they corresponded to rare unique sequences (< 9 reads) and 11,445 (0.2%) of reads because they were chimeric. The remaining reads represented a total of 5751 unique sequences and 5,835,359 reads.

The abundance table produced was next filtered using positive, negative controls and PCR triplicates.

Filtering cross-contaminations using threshold T_{CC}. We observed between 0 and 6,230 reads (total: 34,586; mean: 629 reads; SD: 1289) in the negative controls (NC; *n* = 58). In these NC, 90% of the reads represented 11 unique sequences, and 38% belonging to a human haplotype that was detected with a maximum of 3775 reads in the most contaminated negative control (NC_{PCR}). T_{CC} thresholds ranged between 0 and 3775 reads, depending on the unique sequence considered. After filtering the dataset using these T_{CC} thresholds, we kept 5,704,150 reads representing 5697 unique sequences (Table 2).

Filtering false-assignments of reads using threshold T_{FA}. The 'alien' positive control (PC_{alien}) produced a total of 30,179 reads among which 30,111 were assigned to the nine independent PCRs performed on this DNA. The other 68 reads, i.e. 0.23%, were mis-assigned to 52 other samples, with a maximum of 13 reads observed for a given bat faecal sample. The maximum false assignment rate R_{FA} was therefore equal to 0.043%, and T_{FA} varied between 0 and 244 reads depending on the unique sequence considered. After filtering, the result table included 5697 unique sequences and 5,684,166 reads. Note that T_{FA} excluded reads but not taxa. (Table 2).

Filtering inconsistent results using PCR triplicates. 54.4% of the occurrences (*i.e.* cells showing at least one read in the abundance table) were not triplicated and were removed. Among these inconsistent results, 74.8% were positive for only one of the three PCR triplicates. The remaining reads represented 2636 unique sequences and 5,172,708 reads (Table 2).

Finally, for each sample and unique sequence, the reads of the triplicated PCRs were summed in the abundance table. The 21 bat samples that did not include any read after data filtering were discarded from the dataset.

Comparison of DNA extraction kits

The six different DNA extraction kits tested differed in their performance levels (Table 3). The mean number of reads produced per PCR replicate ranged between 2901 (QIAamp Fast DNA Stool) and 7300 (NucleoSpin Soil) depending on the kit considered. This variation resulted partly from PCR failures whose rate could reach 14% (QIAamp Fast DNA Stool and EZ-10 96 DNA). Successful bat sequencing was observed in 84% of the faecal samples analysed with the NucleoSpin 8 Plant II kit but it only reached 67% to 72% with the five other kits. Successful prey sequencing was observed for 79% to 89% of all faecal samples analysed, with higher success observed with the NucleoSpin soil (89%) and the NucleoSpin 8 Plant II kits (88%).

Mock community analyses

We first confirmed that the arthropod DNA extracts sequenced independently with the MiSeq run produced similar identifications than Sanger sequences when considering the 19 most abundant unique MiSeq sequences (Table 1). These 19 unique sequences were 100% identical to the reference Sanger sequences (Fasta S1). They represented 79% of the reads for these individual PCRs, confirming the high quality of the MiSeq sequencing. BOLD identification tool enabled to identify specifically 14 of the 19 COI minibarcode sequences. *Forficula lesnei* (MC_1) and *Acorypha* sp. (MC_2) were not identified using the public databases BOLD and GENBANK because they were not referenced. Concerning *Monochamus sutor*, *Chrysopa perla* (MC_1) and *Bactrocera dorsalis* (MC_2), we obtained equivalent multi-affiliations with two to five candidate taxa matching with phylogenetically close species (*Monochamus sutor* or *M. sartor*; *Chrysopa perla* or *C. intima*; *Bactrocera dorsalis*, *B. invadens*, *B. philippinensis*, *B. musae* or *B. cacuminata*).

Other non-expected unique sequences were detected at low frequencies (mean: 0.46%, min.: 0.02%, max.: 17.12%) and corresponded to heteroplasmy, pseudogenes (NUMTs: Nuclear insertions of mitochondrial sequences) or to parasitoid sequences. Indeed, in *Forficula lesnei* sample, the tachinid parasitoid *Triarthria setipennis* was detected (Table 1).

We next analysed results of both mock communities. MC_1 sequencing revealed 11 of the 12 insect expected sequences (Table 1). Frequencies of reads varied from 0.4% to 30.1% (expected frequencies 8.3%) while genomic DNA extracts were mixed in equimolar concentrations, therefore revealing biases in PCR amplification. *Protaetia morio* was not detected after data filtering. Insights into the raw dataset showed that some reads were obtained in all three PCRs but with numbers below T_{CC} threshold (T_{CC} = 5 reads for this unique sequence). Sequences of the parasitoid fly *Triarthria setipennis* were detected at low frequency (0.2%). The seven insect species included in MC_2 were detected, with frequencies ranging from 1.5% to 27.3% (expected frequencies 14.3%). For both mock communities, chimeric reads were detected visually despite the filtering processes using UCHIME program, with frequency levels reaching 3.4% and 5.1%, for MC_1 and MC_2 respectively.

Taxonomic identification of bats and their preys

Up to now, less than half of COI sequences deposited in BOLD Systems are made public (Species Level Barcode Records in January 2017: 2,697,359 Sequences/175,125 Species; Public Record Barcode Database: 1,018,338 Sequences/85,514 Species). We therefore had to perform taxonomic assignment using the identification tool implemented in BOLD Systems. We analysed the 1318 most abundant unique sequences of the whole dataset, represented by more than 100 reads and equivalent to 99.1% of all remaining reads.

Further analyses revealed that a majority of the remaining 1318 rare sequences (less than 100 reads) could not be assigned clearly to any taxa. They mainly are chimeric sequences or pseudogenes that had not been removed during the filtering process.

The analysis of the 336 bat faecal samples in PCR triplicates led to the detection of 47 unique sequences corresponding to bat species (1,974,394 reads) and 925 unique sequences belonging to the phylum Arthropoda (1,619,773 reads). Among these latter, 654 unique sequences were assigned with similarity level higher than 97% in BOLD (1,305,633 reads). Finally, 35 unique sequences could not be assigned to any taxa either in BOLD or GENBANK (80,977 reads) (Fig. 2 and Fig. S1). Within samples, the proportion of reads between bats and arthropods was quite balanced, except for *Myotis nattereri*, *Myotis mystacinus* and *Myotis alcathoe* for which lower frequencies of reads were observed (Fig. S1).

Surprisingly, other non-expected taxa were also detected, including 73 unique sequences (133,423 reads) attributed to nematodes (39,186 reads), plants (*Magnoliophyta* : 20,197 reads ; *Bryophyta* : 8247 reads), gastropods (19,915 reads), algae or fungi (*Heterokonta* : 3367 reads ; *Ochrophyta* : 460 reads ; *Oomycota* : 1656 reads ; *Rhodophyta* : 979 reads ; *Zygomycota* : 12,510 reads), rotifers (11,925 reads), targigrades (462 reads), birds (285 reads) as well as mammals (*Felis sp.* : 8077 reads ; *Bos taurus* : 3826 reads ; *Glis glis* : 1658 reads ; *Ovis aries* : 363 reads ; *Mus sp.* : 310 reads). The 1232 remaining rare unique sequences (32,868 reads) corresponded to 0.7% of the 4,932,226 reads in bat samples, and were considered as unclassified (Fig. 2, Fig. S1).

Comparison of field and molecular identification of bat species

We found a congruent taxonomic identification of bat species between molecular and morphological analyses for 238 out of the 336 faecal samples analysed (70.8%, see Fig. 3). Note that for two species sampled, *Myotis myotis* and *Eptesicus serotinus,* the COI minibarcode provided equivalent assignment to two species, *Myotis myotis* and *Myotis blythii* in one hand, and *Eptesicus serotinus* and *Eptesicus nilssonii* in other hand.

The remaining samples mainly resulted in amplification failures for at least one PCR replicate (72 samples, i.e. 21.5%) for one (5.4%), two (5.4%) or the three (10.7%) PCR triplicates performed for each sample (Fig. 3). They mostly concerned *Myotis nattereri* (12 failures over 19 samples tested) and *Rhinolophus ferrumequinum* (24 failures over 60 samples tested). A mismatch (T/C) in the reverse primer could be at the origin of these higher rates of amplification failure for these bat species. Finally, 17 samples provided ambiguous molecular results with two bat species detected for one pellet (Fig. 3). Ten samples provided incongruent taxonomic identification with regard to bat morphology.

Diet composition

In further diet analyses, we considered the 268 bat faecal samples for which we had a strictly congruent taxonomic identification between morphological and molecular analyses based on the results of one, two or three PCR replicates (Fig. 3). We removed samples for which no relevant prey data could be analysed, including 20 samples for which only bat sequences were recovered, six samples with only unclassified sequences, eight samples with only sequences of nematodes, plants, fungi and/or algae, and 18 samples for which arthropod sequences were recovered, but with levels of similarity that were too low (<97%) to provide a reliable assignation to a precise taxa. Altogether, diet compositions were described on 216 bat faecal samples, corresponding to 16 bat

species (Fig. 4). We detected 551 arthropod unique sequences and we identified 18 orders, 117 family, 282 genus and 290 species. The detailed results of diet per bat species are provided in Table S2.

NMDS analyses were realized independently for two sites and showed high variability in diet composition between the bat species analysed. The dataset included *Barbastella barbastellus* (n = 15), *Myotis emarginatus* (n = 17), *Pipistrellus pipistrellus* (n = 23), *Myotis bechsteinii* (n = 26), *Miniopterus schreibersii* (n = 27), *Myotis daubentonii* (n = 29) and *Rhinolophus ferrumequinum* (n = 30). Ordination plots revealed that bat individuals were distributed along both axes, and that bat species slightly clustered although some overlap was also detected. In Jonzac locality, the five bat species of the swarming site formed three distinct clusters, with *Myotis emarginatus* differing from the other bat species due to the high consumption of *Aranea* and *Myotis daubentonii* differing from the other species due to the important consumption of *Diptera* (Fig. 5a). In Saint-Bonnet-sur-Gironde, *Myotis emarginatus* formed a distinct cluster from *Miniopterus schreibersii* and *Rhinolophus ferrumequinum* (Fig. 5b). These results have to be taken cautiously as stress values were higher than 0.2 (0.03 and 0.08 respectively for Jonzac).

Based on GLM, we found significant effects of bat species (p = 0.001) and sites (p = 0.007) on bat diet richness of prey taxa. The diet richness of *Myotis bechsteinii* was significantly lower than the one of *Pipistrellus pipistrellus* ($p < 10^{-4}$) or *Miniopterus schreibersii* (p = 0.025). Diet richness was also lower for bats sampled in the site Vilhonneur than in St-Bonnet-sur-Gironde (p = 0.005).

Discussion

Importance of data filtering and controls

Assessing diet variability between individuals and populations is of main importance to better understand the biology of species, here bats. Producing individual dietary data by DNA metabarcoding requires filtering procedures of prey occurrence based on negative, positive controls and technical replicates. Indeed, it is known that metabarcoding approaches lead to a non-negligible rate of sequence-to-sample misidentification that may result in a high proportion of false positives (Esling et al., 2015 ; Kircher et al., 2012 ; Schnell et al., 2015). Although the procedures limiting false positive results have been well described (Ficetola et al., 2016), few empirical studies include them in their methodological procedures (but see in an epidemiological context, Galan et al., 2016). Although negative controls are often included during DNA extraction and PCRs, they are usually not sequenced and only checked using gel electrophoresis. This procedure is not satisfactory as most contaminating sequences cannot be visually detected. In particular, cross-contaminations of index/tags, tags jumps (Schnell et al., 2015) or polyclonal / mixed clusters during Illumina HTS (Kircher et al., 2012) that may lead to the mis-assignments of reads. The positive control PC_{alien} proposed here enabled to estimate these read mis-assignments for the whole run (0.23%) or for a given sample (up to 0.043%). Finally, mock communities (MC) are scarcely used in animal metabarcoding (but see Pinol et al., 2015) whereas it is widely included in microbial studies to empirically assess the efficiency and biases of both molecular designs and bioinformatic pipelines.

Here, we proposed a set of filtering and validation procedures based on negative, positive controls and independent PCR triplicates. T_{CC} and T_{FA} thresholds were estimated from controls and enabled to remove false positives (respectively due to pre-sequencing contaminations and polyclonal/mixed clusters during sequencing) while triplicates permitted to remove random sequencing errors generating non-repeatable occurrences

(Robasky et al., 2014). Our results showed that applying T_{CC} and T_{FA} removed relatively few reads and unique sequences compared to the triplicate validation procedure. Hence the rates of laboratory contaminations or mis-assignments during HTS seemed to be low while the proportion of non-repeatable unique sequences seemed to be high. The reasons might be methodological (*e.g.* PCR chimera, sequencing errors or PCR drop-out) or biological (presence of NUMTs at low frequencies, low biomass preys, traces of ancient meals). Finally, we recommended to include arthropod artificial communities in metabarcoding studies to assess the sensitivity of the method and the amplification biases.

Methodological framework to avoid methodological biases

Fieldwork remains a crucial step for diet analyses, even with such molecular approaches. The way faecal pellets are collected may lead to cross-individual contaminations, as revealed in our study by the detection of bat sequences corresponding to the wrong species or a mix of two species for a given faecal sample. Considering invasive sampling, cotton holding bags in which individuals are kept before morphological analyses must be carefully checked to avoid the collection of faecal samples belonging to successively captured bats. Faecal pellets should also be carefully handled and stored to avoid contaminations. These points reinforce the importance of performing species bat identification for each pellet to prevent mis-assignment of a diet and to an individual of the wrong species. In the case of non-invasive sampling, it is highly recommended to use clean supports (e.g. unused polythene sheeting, Hope et al., 2014) and single-use instruments to collect the pellets. Reducing time-delay between bat faeces release and collection might also be important to avoid DNA degradation and

cross-contaminations due to urine from different species of bats or coprophagous arthropods for example.

Further storage conditions of faecal samples are of main importance to guarantee DNA integrity and limit the proliferation of micro-organisms. Some amplification failures observed in our study were hence likely to result from the storage of bat faecal samples at room temperature in empty tubes during several months. Samples should be frozen at -20°C at least, or stored in an appropriate storage buffer to prevent DNA degradation (Renshaw, Olds, Jerde, McVeigh, & Lodge, 2015). We also showed that extraction methods may influence the success of metabarcoding studies. Our results evidenced that NucleoSpin 8 Plant II kit provided the best result in terms of host sequencing success and the second best results with regard to the number of reads produced per PCR, PCR failure, and prey sequencing success. It therefore seems to be the best compromise between cost per extraction, throughput (96-well format) and quality of the DNA purification for metabarcoding applications.

In addition, we have evidenced PCR amplification biases for particular prey and bat taxa that lead to a high variability in sequencing depth and even to the amplification failure for particular species. For example, a mismatch in the reverse primer used to amplify the COI minibarcode probably explained the lower amplification success observed for *Myotis nattereri*, *Myotis mystacinus* and *Rhinolophus hipposideros* samples. If necessary, one could add a degenerated base at position 21 of this reverse primer (*i.e.* EPT-long-univR-MiSeq: 5'-ACTATAAAARAAAYTATDAY**R**AADGCRTG-3') to increase the proportion of reads corresponding to bat species. The analyses of mock communities also revealed amplification biases with regard to arthropod species, as previously described in other empirical studies (e.g. Pinol et al. 2015). One species (*Protaetia morio*) over the 12 included in MC_1 could not be detected although such amplification

bias was not observed when sequencing the arthropod species of this community individually. This result was probably due to mismatches between PCR primers and arthropod species DNA, and to further primer annealing competition during DNA amplification of the community. It is noteworthy that designing COI universal primers generating no PCR amplification biases might not be achievable as shown by Deagle, Jarman, Coissac, Pompanon, and Taberlet (2014). Therefore they proposed the use of ribosomal rRNA, either mitochondrial (16S) or nuclear (12S), with conserved flanking regions among taxonomical distant species. Although this proposal should reduce the probability of primer-template mismatches (and therefore amplification biases among taxa), it is yet not suitable for arthropod metabarcoding approaches due to the absence of public databases similar to BOLD (*i.e.* curated database with reference sequences linked to taxonomically verified voucher specimens). In conclusion, COI minibarcodes still remain an imperfect, but 'not so bad' solution for taxonomic identification when this latter requires to reach the species level. Our approach is compatible with a multi-primer approach that could ensure to recover particular species of interest for studies where the design of a single couple of universal primers failed (Miya et al., 2015).

Prey-bat simultaneous identification and taxonomic resolution

In addition to the methodological framework provided, the originality of our approach also resided in the simultaneous identification of prey and bat species. Despite the low quality of our faecal samples due to inappropriate storage conditions, we obtained a specific identification for 74% of the bats studied, including only 3% of incongruent results between the molecular and field identifications due to remaining faecal pellets in the cotton holding bags (see above). Among the 26% unidentified, 21% were due to bat sequencing failure of at least one PCR replicate, 5% corresponded to the detection of

several bat species. The failure rate associated with our molecular approach was therefore congruent with what is observed in traditional Sanger sequencing of bats (e.g. 19% reported in Hope et al., 2014), and could be easily improved with the use of appropriate storage conditions and DNA extraction method. Our approach is thus relevant for bat species identification too.

The simultaneous identification of predators and preys is generally avoided in diet analyses as it may induce biases in the pool of sequences produced (Pompanon et al., 2012). In particular, high success of predator amplification will reduce prey amplification, what will in turn affect the sensibility of diet analyses. Here, we reported well-balanced proportions of reads for bats and their preys, with a mean of 43.3% of bat reads whereas other studies reported higher proportions (91.6% for the leopard cat in Shehzad et al., 2012). The lower proportion of bat reads observed in our study may result from a lower proportion of the predator DNA in the faecal pellets analysed or to a lower primer specificity to amplify target DNA from bats. These latter enabled an important sequencing depth for arthropods that guarantees the high sensitivity for prey detection.

Our approach provided a high level of taxonomic resolution for bats and their preys as evidenced by the congruent identifications obtained using the 658bp Sanger sequences and the 133bp minibarcode for the arthropod mock community. Particularly, we confirmed the possibility to discriminate morphologically close insect species including the pine processionary moth complex *Thaumetopoea pityocampa*/*T. wilkinsoni* (Kerdelhue et al., 2009), the longhorn beetle *Monochamus galloprovincialis* and its sister species *M. sutor* (Haran, Koutroumpa, Magnoux, Roques, & Roux, 2015) or the green lacewing *Chrysopa perla* / *C. formosa* (Bozsik, 1992). Such taxonomic resolution is highly important when dealing with arthropod pests or arthropod species involved in

biological pest control. In particular cases (Mayer & von Helversen, 2001), knowledge on geographic distribution may help deciphering the most likely taxa in presence (e.g. *Eptesicus serotinus* and *Eptesicus nilssonii*, or *Myotis myotis* and *Myotis blythii*).

We have also evidenced some limitations with regard to public sequence databases. Two arthropod species included in the mock communities could not be identified as they were not included in BOLD and GENBANK databases. We have also emphasized some errors that may have consequences for further identification. For example, one sequence of *Pipistrellus pipistrellus* included in BOLD is mis-assigned to *Pipistrellus kuhlii* (GENBANK Accession JX008080 / BOLD Sequence ID: SKBPA621-11.COI-5P). This mistake made it impossible to distinguish both species. Although recently reported (Shen, Chen, & Murphy, 2013), it has not yet been cleaned. It seems that such errors are quite frequent in GENBANK (Shen et al., 2013), and unfortunately BOLD database does not appear to be spared. Completeness and reliability of public databases are therefore still a main pitfall in metabarcoding studies.

Dietary composition for 16 bat species in Western France

The combination of HTS and filtering procedures described here provided a detailed diet characterization of the 16 bat species sampled in Western France. At the order level, our results were congruent with previous knowledge of preys consumed detected using morphological analyses. For example, the diet of *Myotis daubentonii* and *Pipistrellus pipistrellus* is known to be dominated by Diptera (e.g. Arlettaz, Godat, & Meyer, 2000; Vesterinen, Lilley, Laine, & Wahlberg, 2013), what was confirmed by our results showing respectively 79% and 74% of faeces samples positive for Diptera. Lepidoptera were also highly frequent in *Pipistrellus pipistrellus* samples, being the second order detected in faecal samples (48% of positive faecal pellets), as described in Arlettaz et al.

(2000). The diet of *Barabastella barbastellus* was dominated by Lepidoptera with 100% positive samples, as previously described in Andreas, Reiter, and Benda (2012). Finally, the lowest dietary richness was found for *Myotis emarginatus*, with sequences of Arenae detected in 100% of faecal samples, what was previously described in Goiti et al. (2011).

Compared to these morphological studies, our approach provided greater details on dietary composition by increasing prey taxonomic resolution and enabling to distinguish between species that can not be recognized based on morphological criteria only, as described above. In addition, our approach enabled to detect unexpected interactions including secondary predation events and gastrointestinal infestations. We reported the presence of snails (*Cepaea hortensis* and *Cepaea nemoralis*) and slugs (*Arion intermedius*) in diets that included Carabidae preys (*Abax parallelepipedus* and *Carabus* sp.) in three *Myotis myotis* samples. Also surprisingly, three *Myotis nattereri* and one *Myotis daubentonii* bat faecal samples contained *Bos taurus* sequences. It is likely that this finding results from traces of bovine animal blood in the gut of the biting house fly *Stomoxys calcitrans* that were also detected as consumed preys in the *M. daubentonii* sample, or traces of bovine animal excrements coming from green bottle fly (*Neomyia cornicina*), crane flies (*Tipula* sp.) or scavenger cokroach (*Ectobius* sp.) for the *M. nattereri* samples. Similarly, grey heron (*Ardea cinerea*) and dormouse (*Glis glis*) sequences found in two *Myotis daubentonii* bat faecal samples may be due to blood traces in biting insects, the common house mosquito *Culex pipiens* and the autumn house-fly *Musca autumnalis* respectively. The fact that prey diet traces can be found in bat faeces can be used to improve our knowledge of trophic relationships. Finally, nematode sequences were emphasized in 25 bat faecal samples, revealing the infestation of these individuals with gastro-intestinal worms. Unfortunately, the COI minibarcode did not enable to provide a specific identification for this phylum.

Ecological importance of molecular diet analyses

Our dataset enabled to reveal the presence in bat diets of 61 pest species (Table S2) among which some are important for agricultural management (*e.g.* the cotton bollworm *Helicoverpa armigera*, the spotted-wing drosophila *Drosophila suzukii* and the pine processionary *Thaumetopoea pityocampa*) or veterinary and Public Health issues (*e.g.* the malaria vector *Anopheles claviger,* the biting house fly *Stomoxys calcitrans*, the common house mosquito *Culex pipiens*). Our results therefore confirm the possibility to use our metabarcoding approach as an indirect tool for "chirosurveillance" without any *a priori* with regard to the pests that need to be surveyed (Maslo et al., 2017). They could help for the design of prevention and control strategies.

This study also demonstrated the capacity of our approach to reveal variation in diet richness and composition within sites, between sites and between species. Until recently, the study of species co-existence and interaction was hampered by an imprecise understanding of exactly what most species actually prey under different environmental conditions. Combining bat-species molecular identification with insect diet analyses will provide a more complete understanding of how bat diet varies along season, life history stage, gender and age, and will permit to answer questions about niche size and niche overlap. The DNA metabarcoding method will also revolutionize our ability to investigate the abiotic and biotic factors, *e.g.,* landscape structure, crop type, and insect abundance, which define the use of agricultural landscapes by insectivorous bats. Understanding these patterns is crucial for identifying species requirements and predicting their responses to human disturbance and thus for developing appropriate bat conservation strategies.

Conclusion

The DNA metabarcoding approach described here enables the simultaneous identification of bat species and their arthropod diets from faeces, for several hundreds of faecal samples analysed at once. This strategy involves twice less steps than usually required in other metabarcoding studies and reduces the probability to mis-assigned preys to the wrong bat species. The two-step PCR protocol proposed here makes easier the construction of libraries, multiplexing and HTS, at a reduced cost (about 8€ per faecal pellet). Our study also includes several controls during the lab procedures associated to a bioinformatic strategy that enables to filter data in a way that limits the risk of false positive and that guarantees high confidence results for both prey occurrence and bat species assignment. This study therefore provides a rapid, resolutive and cost-effective screening tool for addressing 'chirosurveillance' application or evolutionary ecological issues in particular in the context of bat conservation biology. It may be easily adapted for use in other vertebrate insectivores.

Acknowledgements

First of all, we are indebted to all of our collaborators in the naturalist associations who helped us to gather and generate this comprehensive bat sampling: Matthieu Dorfiac (Charente Nature), Anthony Le Guen (Deux-Sèvres Nature Environnement), Virginie Barret and Philippe Jourde (LPO France), Sébastien Roué (Groupe Chiroptères Aquitaine) and all the numerous volunteers that contributed to field work.

We thank Jean-Claude Streito, Guenaelle Genson, Antoine Foucard, Laure Benoit, Carole Kerdelhué and Laure Sauné for the insect DNA provided to constitute the mock communities. Many thanks to Emese Meglecz, Vincent Dubut and Emmanuel Corse for the constructive discussions and sharing about the metabarcoding analyses. Data used in this work were partly produced through the genotyping and sequencing facilities of ISEM (Institut des Sciences de l'Evolution-Montpellier) and LabEx Centre Méditerranéen Environnement Biodiversité. Analyses were performed on the CBGP HPC computational platform thanks to Alexandre Dehne-Garcia. This work was funded by the LabEx ECOFECT and the internal funding from the CBGP laboratory. Orianne Tournayre PhD is funded by the LabEx CeMEB. Part of this work (JBP, DP) was performed within the framework of the LabEx ECOFECT (ANR-11-LABX-0048) of Université de Lyon and the Poitou-Charentes Nature Feder project « *Grand Rhinolophe et trame verte bocagère : étude des facteurs environnementaux influant sur la dynamique des populations* » (ML, JBP, DP).

References

Andreas, M., Reiter, A., & Benda, P. (2012). Prey selection and seasonal diet changes in the western barbastelle bat (*Barbastella barbastellus*). *Acta Chiropterologica, 14*, 81-92. doi:10.3161/150811012x654295

Arlettaz, R., Godat, S., & Meyer, H. (2000). Competition for food by expanding pipistrelle bat populations (*Pipistrellus pipistrellus*) might contribute to the decline of lesser

horseshoe bats (*Rhinolophus hipposideros*). *Biological Conservation, 93*, 55-60. doi:10.1016/s0006-3207(99)00112-3

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS one, 2*. doi:10.1371/journal.pone.0000197

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., . . . de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution, 29*, 358-367. doi:10.1016/j.tree.2014.04.003

Bohmann, K., Monadjem, A., Noer, C. L., Rasmussen, M., Zeale, M. R. K., Clare, E., . . . Gilbert, M. T. P. (2011). Molecular diet analysis of two african free-tailed bats (Molossidae) using high throughput sequencing. *PloS one, 6*. doi:10.1371/journal.pone.0021441

Bozsik, A. (1992). Natural adult food of some important Chrysopa species (Planipennia: Chrysopidae). *Acta Phytopathologica et Entomologica, 27*, 141–146.

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., & Geigl, E. M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PloS one, 5*. doi:10.1371/journal.pone.0013042

Clare, E. L. (2014). Molecular detection of trophic interactions: emerging trends, distinct advantages, significant considerations and conservation applications. *Evolutionary Applications, 7*, 1144-1157. doi:10.1111/eva.12225

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., . . . Hall, N. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics, 17*. doi:10.1186/s12864-015-2194-9

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters, 10*. doi:10.1098/rsbl.2014.0562

Deiner, K., Walser, J. C., Machler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation, 183*, 53-63. doi:10.1016/j.biocon.2014.11.018

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*, 2194-2200. doi:10.1093/bioinformatics/btr381

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research, 43*, 2513-2524. doi:10.1093/nar/gkv107

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome, 2*. doi:10.1186/2049-2618-2-6

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources, 16*, 604-607. doi:10.1111/1755-0998.12508

Galan, M., Razzauti, M., Bard, E., Brouat, C., Charbonnel, N., Dehne-Garcia, A., . . . Cosson, J. F. (2016). 16S metagenomics for epidemiological survey of bacteria in wildlife. *mSystem, 1*, e00032-00016. doi:10.1128/mSystems.00032-16

Gillet, F., Tiouchichine, M. L., Galan, M., Blanc, F., Nemoz, M., Aulagnier, S., & Michaux, J. R. (2015). A new method to identify the endangered Pyrenean desman (*Galemys*

pyrenaicus) and to study its diet, using next generation sequencing from faeces. *Mammalian Biology, 80*, 505-509. doi:10.1016/j.mambio.2015.08.002

Goiti, U., Aihartza, J., Guiu, M., Salsamendi, E., Almenar, D., Napal, M., & Garin, I. (2011). Geoffroy's bat, *Myotis emarginatus*, preys preferentially on spiders in multistratified dense habitats: a study of foraging bats in the Mediterranean. *Folia Zoologica, 60*, 17-24.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., . . . Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution, 7*, 1299-1307. doi:10.1111/2041-210x.12595

Gonsalves, L., Law, B., Webb, C., & Monamy, V. (2013). Foraging ranges of insectivorous bats shift relative to changes in mosquito abundance. *PloS one, 8*. doi:10.1371/journal.pone.0064081

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental barcoding: a next-generationsequencing approach for biomonitoring applications using river benthos. *PloS one, 6*, e17497. doi:10.1371/journal.pone.0017497

Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes, 6*, 959-964. doi:10.1111/j.1471-8286.2006.01470.x

Haran, J., Koutroumpa, F., Magnoux, E., Roques, A., & Roux, G. (2015). Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research, 53*, 109-115. doi:10.1111/jzs.12095

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, B, 270*, 313-321. doi:10.1098/rspb.2002.2218

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 14812-14817. doi:10.1073/pnas.0406166101

Hope, P. R., Bohmann, K., Gilbert, M. T. P., Zepeda-Mendoza, M. L., Razgour, O., & Jones, G. (2014). Second generation sequencing and morphological faecal analysis reveal unexpected foraging behaviour by *Myotis nattereri* (Chiroptera, Vespertilionidae) in winter. *Frontiers in zoology, 11*. doi:10.1186/1742-9994-11-39

Iwanowicz, D. D., Vandergast, A. G., Cornman, R. S., Adams, C. R., Kohn, J. R., Fisher, R. N., & Brehme, C. S. (2016). Metabarcoding of fecal samples to determine herbivore diets: a case study of the endangered pacific pocket mouse. *PloS one, 11*. doi:10.1371/journal.pone.0165366

Kerdelhue, C., Zane, L., Simonato, M., Salvato, P., Rousselet, J., Roques, A., & Battisti, A. (2009). Quaternary history and contemporary patterns in a currently expanding species. *Bmc Evolutionary Biology, 9*. doi:10.1186/1471-2148-9-220

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research, 40*. doi:10.1093/nar/gkr771

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology, 79*, 5112-5120. doi:10.1128/aem.01043-13

Lam, M. M. Y., Martin-Creuzburg, D., Rothhaupt, K. O., Safi, K., Yohannes, E., & Salvarina, I. (2013). Tracking diet preferences of bats using stable isotope and fatty acid signatures of faeces. *PloS one, 8*. doi:10.1371/journal.pone.0083452

Maine, J. J., & Boyles, J. G. (2015). Bats initiate vital agroecological interactions in corn. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 12438-12443. doi:10.1073/pnas.1505413112

Maslo, B., Valentin, R., Leu, K., Kerwin, K., Hamilton, G. C., Bevan, A., . . . Fonseca, D. M. (2017). Chirosurveillance: the use of native bats to detect invasive agricultural pests. *PloS one, 12*. doi:10.1371/journal.pone.0173321

Mayer, F., & von Helversen, O. (2001). Cryptic diversity in European bats. *Proceedings of the Royal Society of London, B, 268*, 1825-1832. doi:10.1098/rspb.2001.1744

McCracken, G. F., Westbrook, J. K., Brown, V. A., Eldridge, M., Federico, P., & Kunz, T. H. (2012). Bats track and exploit changes in insect pest populations. *PloS one, 7*. doi:10.1371/journal.pone.0043839

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., . . . Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Proceedings of the Royal Society of London, B, 2*, 150088. doi:10.1098/rsos.150088

Pinol, J., Mir, G., Gomez-Polo, P., & Agusti, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources, 15*, 819-830. doi:10.1111/1755-0998.12355

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology, 21*, 1931-1950. doi:10.1111/j.1365-294X.2011.05403.x

Razgour, O., Clare, E. L., Zeale, M. R. K., Hanmer, J., Schnell, I. B., Rasmussen, M., . . . Jones, G. (2011). High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and evolution, 1*. doi:10.1002/ece3.49

Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., & Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. *Molecular Ecology Resources, 15*, 168-176. doi:10.1111/1755-0998.12281

Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics, 15*, 56-62. doi:10.1038/nrg3655

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology, 75*, 7537-7541. doi:10.1128/aem.01541-09

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources, 15*, 1289-1303. doi:10.1111/1755-0998.12402

Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., . . . Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology, 21*, 1951-1965. doi:10.1111/j.1365-294X.2011.05424.x

Shen, Y. Y., Chen, X., & Murphy, R. W. (2013). Assessing DNA barcoding as a tool for species identification and data quality control. *PloS one, 8*, e57125. doi:10.1371/journal.pone.0057125

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology, 21*, 2045-2050. doi:10.1111/j.1365-294X.2012.05470.x

Tillmar, A. O., Dell'Amico, B., Welander, J., & Holmlund , G. (2013). A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures. *PloS one, 8*, e83761. doi:10.1371/journal.pone.0083761

Vesterinen, E. J., Lilley, T., Laine, V. N., & Wahlberg, N. (2013). Next generation sequencing of fecal DNA reveals the dietary diversity of the widespread insectivorous predator Daubenton's bat (*Myotis daubentonii*) in Southwestern Finland. *PloS one, 8*. doi:10.1371/journal.pone.0082168

Wright, E. S., & Vetsigian, K. H. (2016). Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics, 17*, 876. doi:10.1186/s12864-016-3217-x

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., . . . Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PloS one, 6*. doi:10.1371/journal.pone.0024570

Author contributions

The study was conceived and designed by M.G., D.P. and N.C. J.-B.P and M.L. supervised the field work. M.G. and J.-B.P carried out the molecular biology procedures and validated the MiSeq data. M.G. contributed to the development of bioinformatics methods and M.G., J.-B.P and E.P. validated taxonomic assignments. D.P. and N.C. coordinated the funding projects (resp. Ecofect and AAP CBGP). M.G., J.-B.P, O.T. and N.C. analysed the data. M.G. and N.C. wrote the manuscript. J.-B.P, O.T., E.P., M.L and D.P. helped to draft and to improve the manuscript. All coauthors read and approved the final manuscript.

Data accessibility

Raw sequence reads (fastq format), raw table of abundance and filtered table of abundance including taxonomic affiliations have been deposited in the Dryad Digital Repository (under embargo during the process of peer review) and are available on request to the corresponding author.

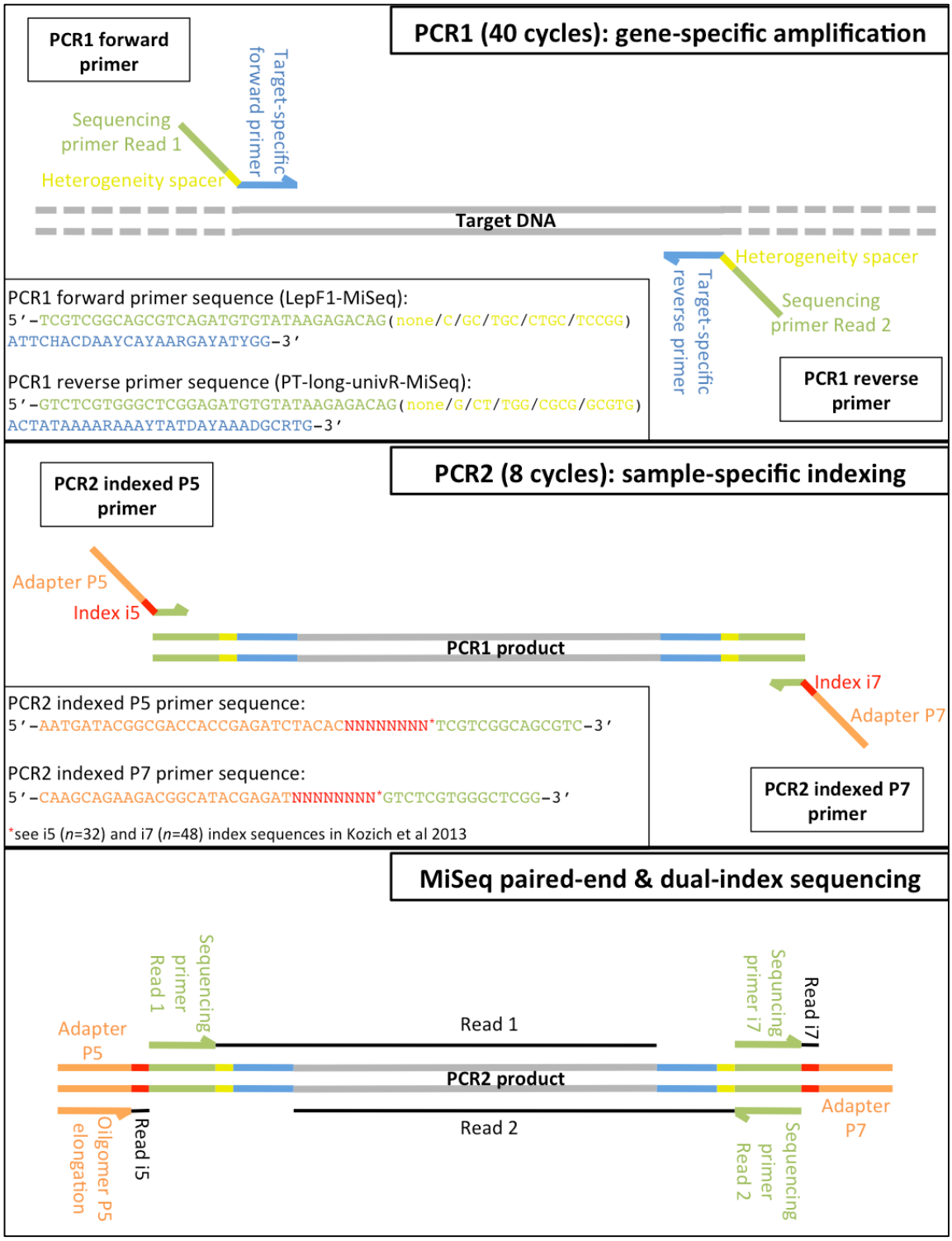


Fig. 1 Schematic description of the library construction using 2-step PCR and MiSeq sequencing.

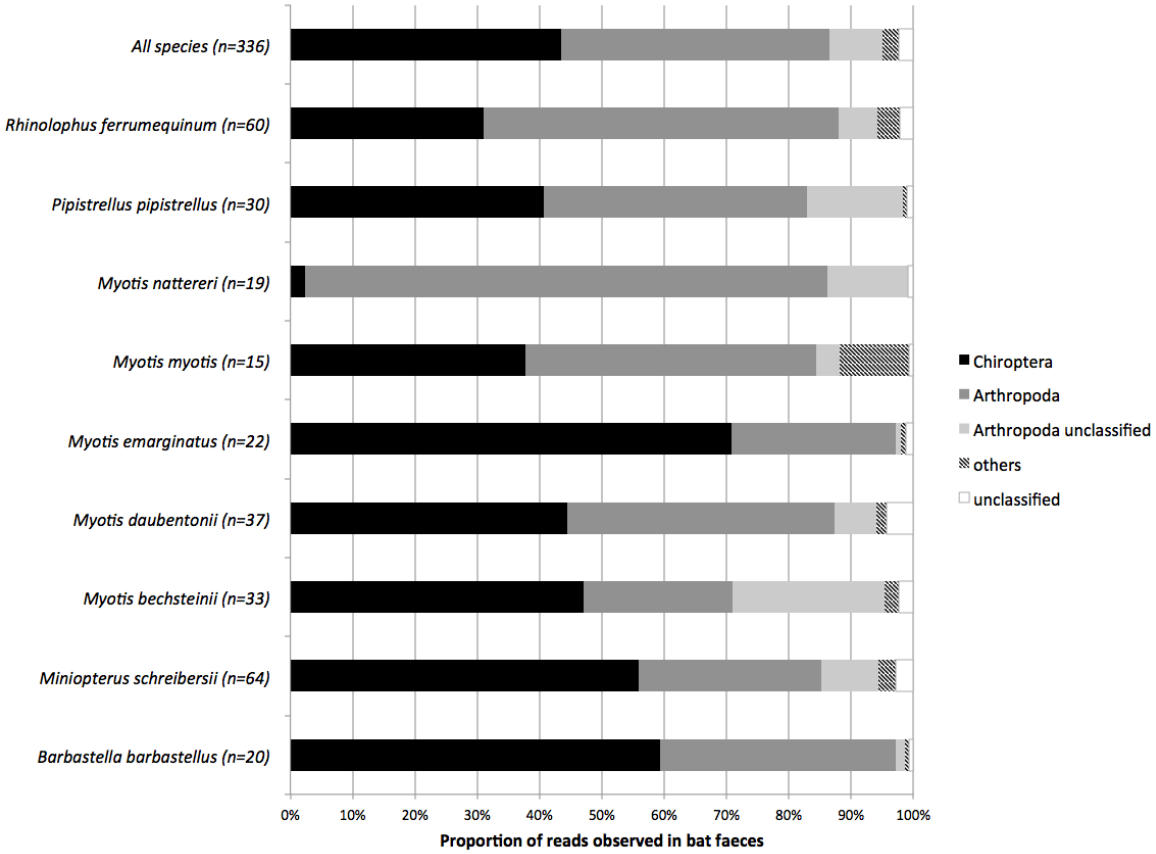


Fig. 2 Proportion and origin of reads obtained from high-throughput sequencing of faecal pellets of different bat species.
Only bat species showing at least 15 individuals are illustrated. Chiroptera reads (black bars) and Arthropoda reads (dark grey bars) correspond to sequences assigned to one or several species using BOLD or GENBANK and identity scores higher than 97%. "Arthropoda unclassified" reads (light grey bars) correspond to sequences assigned to Arthropoda phylum with identity scores lower than 97%. "Other" reads (striped bars) include sequences from putative contaminants (*i.e.* human, cat or fungi), blood meal/coprophagia/necrophagia (others mammalia or birds), secondary preys (*i.e.* Mollusca), parasites (Nematoda) and insect diets (*i.e.* plants). "Unclassified" reads (white bars) correspond to unique sequences with no match in BOLD and GENBANK. The "all species" line provides results from 336 samples including the 16 bat species analysed (i.e. even species with less than 20 individuals). See Figure S1 for detailed results for each of the 16 bat species analysed.

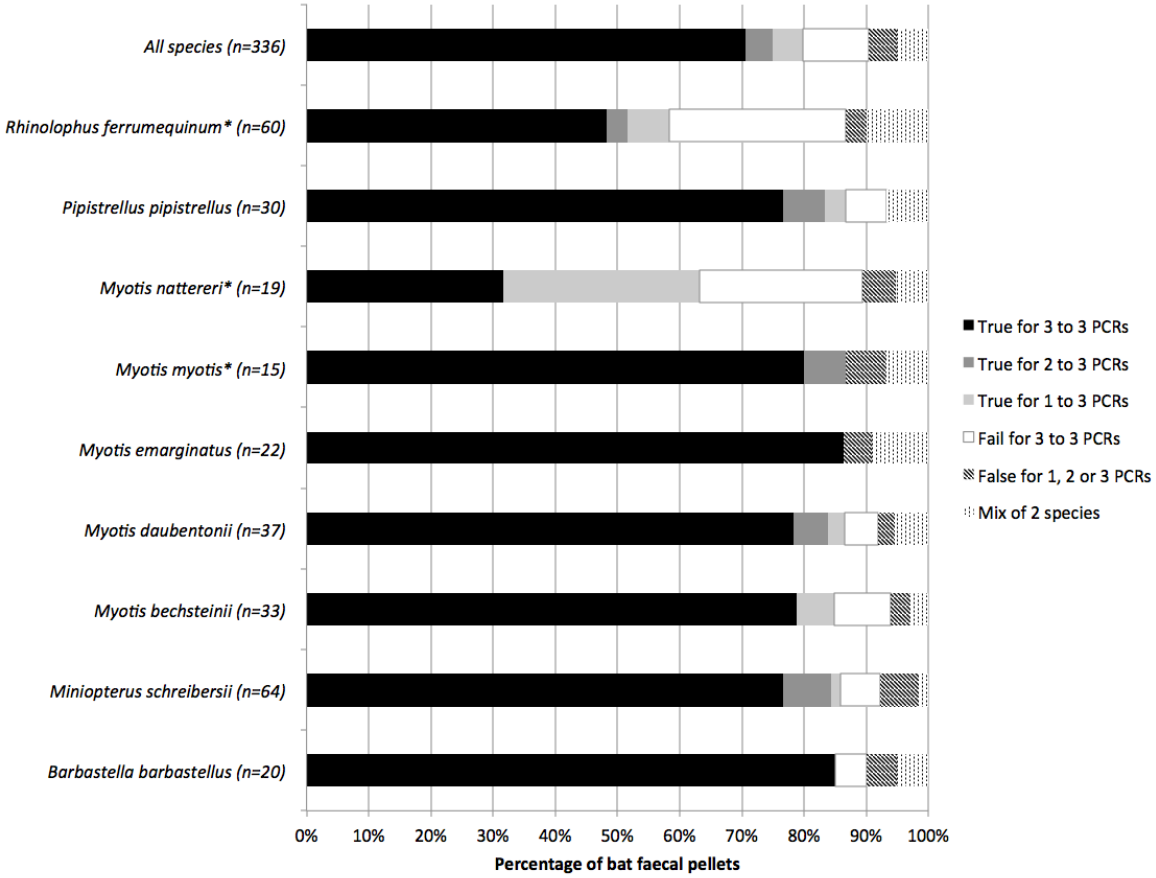


Fig. 3 Proportion of faecal pellets with true or false bat species molecular identification using high-throughput sequencing.
Only bat species showing at least 15 individuals are illustrated. Molecular identifications were compared to the morphological identifications performed during the fieldwork to determine the true or false status. * indicates bats species showing a substitution T-->C that creates a mismatch between the COI target sequence and the reverse PCR primer. This substitution could explain a lower rate of identification success for these species. See Figure S2 for detailed results for each of the 16 bat species analysed.

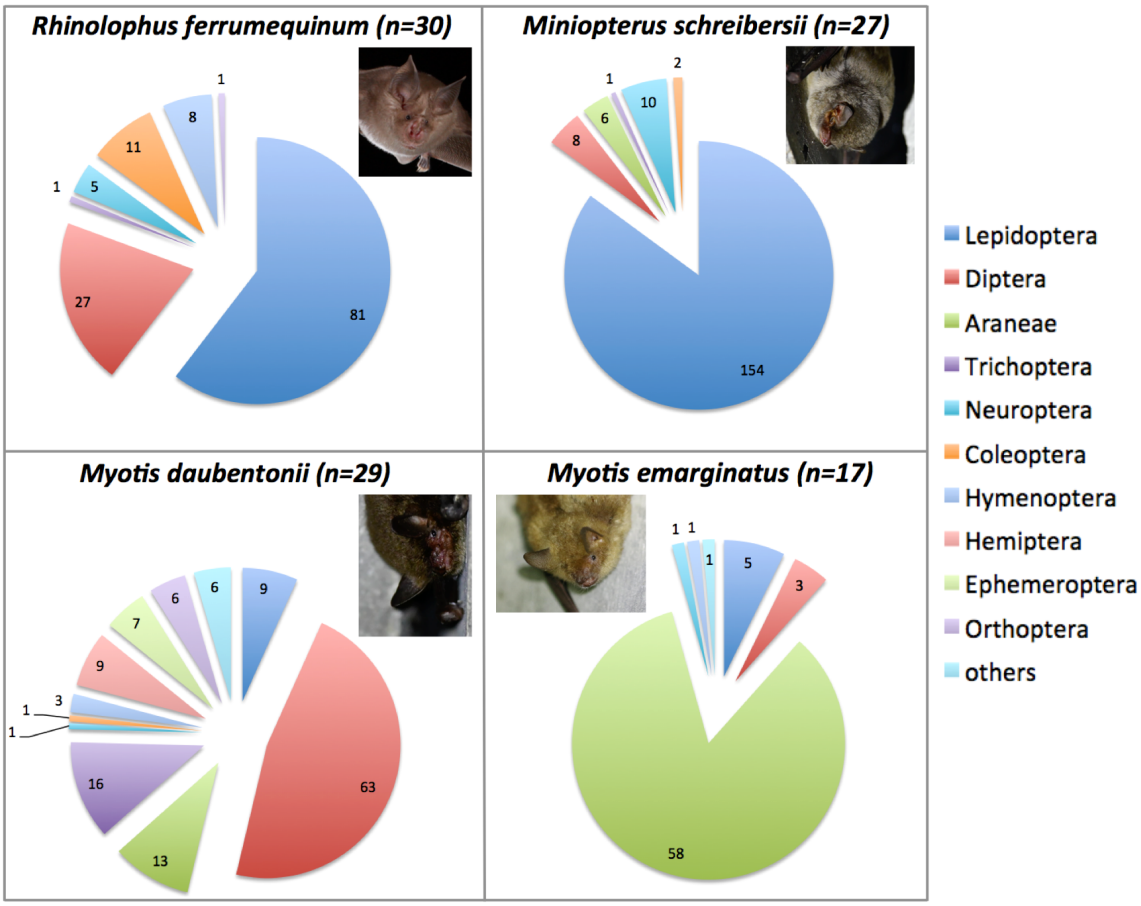
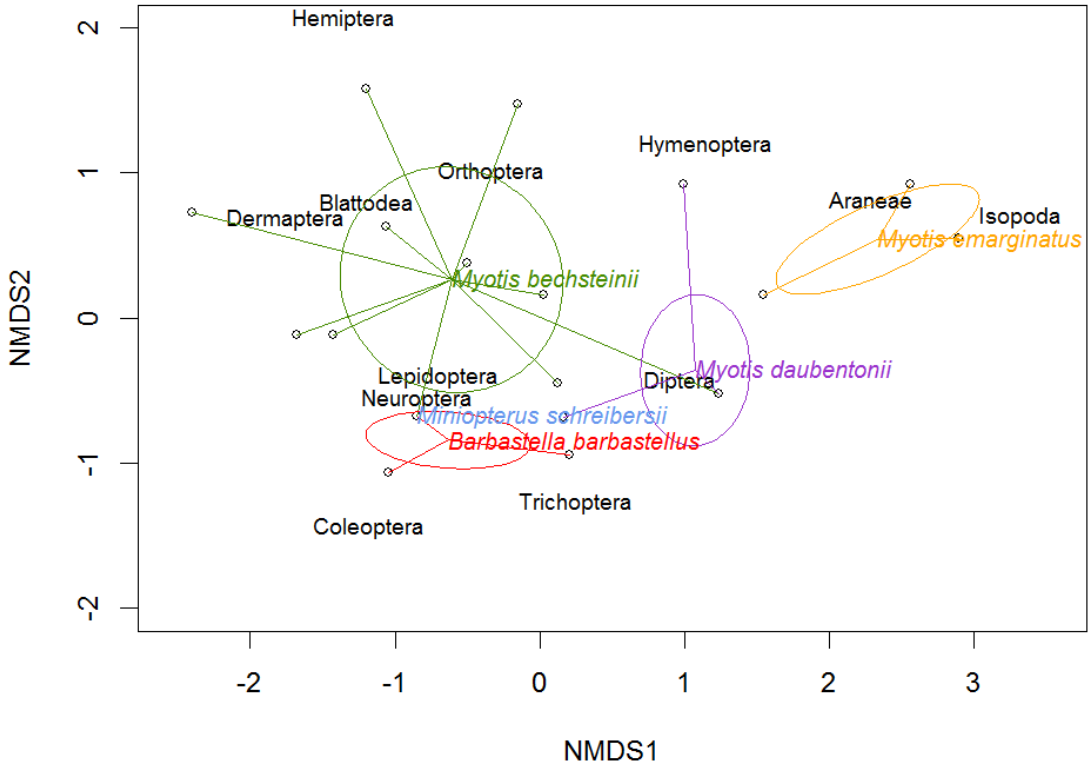


Fig. 4 Number of occurrence of prey taxa within faecal pellets from four bat species. Photos by: Jérémy Dechartre, Matthieu Dorfiac & Maxime Leuchtmann

a)



b)

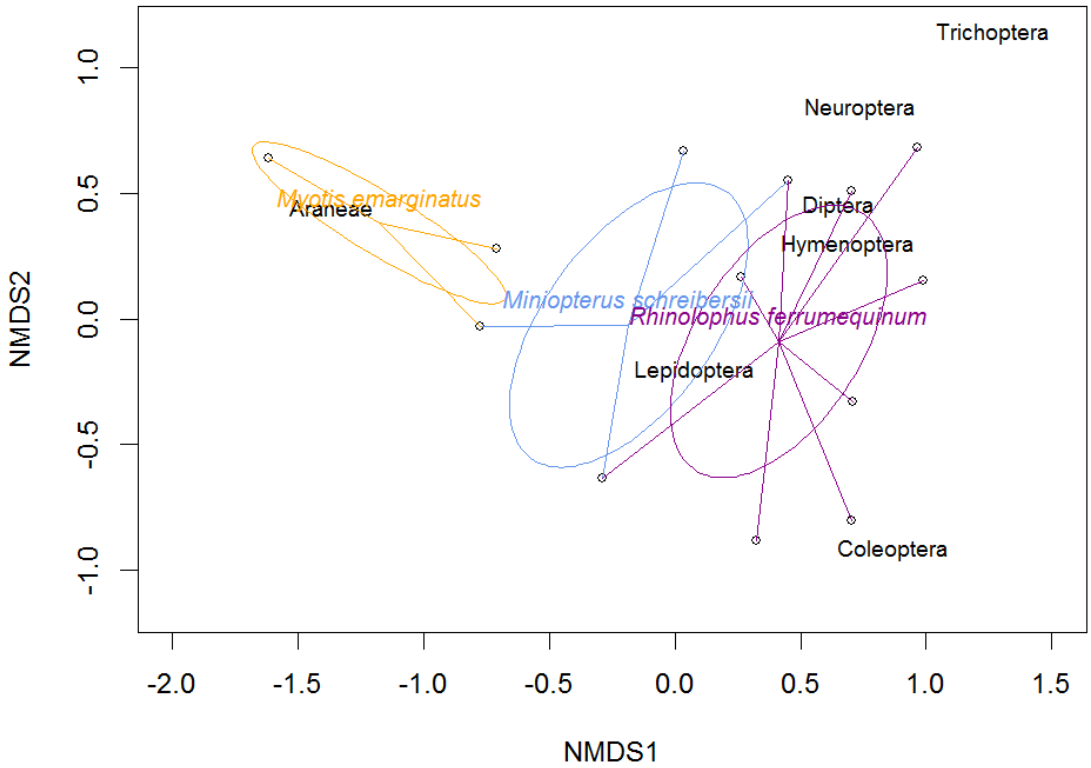


Fig. 5 Nonmetric multidimensional scaling plots representing diet dissimilarity among bat individuals and species. a) Jonzac locality, b) St Bonnet sur Gironde localities.

Supplementary information legends

Table S1 Information about the study samples, the laboratory controls and the technical replicates.

Table S2 List of prey identified within faecal pellets from 16 bat species using high-throughput sequencing. Only results with sequence similarity >97% are kept. We applied a modified version of the criteria described in (Razgour et al., 2011) for the confidence level of the sequence assignments: 1a: match to only one species in BOLD System and >99% sequence similarity; 1b: match to only one species in BOLD System and >98% sequence similarity; 2: several species to the same genus and >98% sequence similarity; 3: several species to different genus of the same family and >98% sequence similarity; 4: several species to different family and/or >97% to <98% sequence similarity. * indicates agronomic pest species. ** indicates pecies of veterinary and medical interest

Fasta S1 Alignment of the Sanger and MiSeq COI sequences for the 19 samples of the mock communities. For the MiSeq sequences, only the most abundant unique sequence for each specimen is included, excepted for *Forficula lesnei* sample where the unique sequence of the parasitoid *Triarthria setipennis* is also included.

Figure S1 Proportion and origin of reads obtains by high-throughput sequencing of faecal pellets from 16 bat species.

Chiroptera and Arthropoda reads correspond to sequence assigned at one or several species in BOLD or GENBANK with an identity >97%. "Arthropoda unclassified" reads correspond to sequence assigned to Arthropoda phylum with an identity <97%. "Others"

reads include sequences from putative contaminants (i.e. human or fungi), blood meal/coprophagia/necrophagia (others mammalia or birds), secondary preys (i.e. Mollusca), parasites (Nematoda) and diet from insects (i.e. plants). "Unclassified" reads correspond to unique sequences with no match in BOLD and GENBANK. The bar "All species" correspond to 336 samples from 16 bat species including species with less of 20 individuals.

Figure S2 Proportion of faecal pellets with true or false bat species molecular identification using high-throughput sequencing.

Molecular identifications were compared to the morphological identifications performed during the fieldwork to determine the true or false status. * indicates bats species showing a substitution T-->C that creates a mismatch between the COI target sequence and the reverse PCR primer. This substitution could explain a lower rate of identification success for these species.

Table 1 Proportion of reads for two mock arthropod communites (MC_1 & MC_2). The species names in the column 'mock community content' are based on the morphological identification realised by experts for each taxa group. The theoretical proportions correspond to the genomic DNA proportion for each species in the equimolar mix of DNA (*i.e.* the pool) before PCR amplification of the mock community. The observed proportions correspond to the mean proportions of the number of reads for the three PCR replicates. The observed proportions in the individual PCRs are calculated after summing the number of reads for all the individual PCRs of each species

	Mock community content	Order	Theoretical proportions (%)	Observed proportions in the triplicated individual PCRs (% ± SD)	Observed proportions in the triplicated pools (% ± SD)	Identification using Bold: top hit(s) & identity (%)
Mock community MC_1	*Monochamus galloprovincialis*	Coleoptera	8.3%	8.2% ± 0.5	30.1% ± 2.1	*Monochamus galloprovincialis* (100%)
	Forficula lesnei	Dermaptera	8.3%	6.0% ± 0.3	14.8% ± 0.5	unclassified (<90%)
	Thaumetopoea pityocampa	Lepidoptera	8.3%	8.4% ± 0.8	9.1% ± 0.7	*Thaumetopoea pityocampa* (100%)
	Athous bicolor	Coleoptera	8.3%	5.5% ± 1.4	7.4% ± 0.8	*Athous bicolor* (100%)
	Forficula auricularia	Dermaptera	8.3%	5.9% ± 0.5	7.1% ± 0.2	*Forficula auricularia* (100%)
	Monochamus sutor	Coleoptera	8.3%	8.2% ± 0.6	7.0% ± 0.4	*Monochamus sutor/sartor* (100%)
	Chrysopa formosa	Neuroptera	8.3%	10.8% ± 0.5	4.3% ± 0.4	*Chrysopa formosa* (100%)
	Chrysopa perla	Neuroptera	8.3%	8.7% ± 0.5	3.6% ± 0.2	*Chrysopa perla/intima* (100%)
	Himacerus mirmicoides	Hemiptera	8.3%	9.4% ± 0.4	2.7% ± 0.6	*Himacerus mirmicoides* (100%)
	Calliptamus barbarus	Orthoptera	8.3%	5.0% ± 0.4	1.7% ± 0.2	*Calliptamus barbarus* (100%)
	Phymata crassipes	Hemiptera	8.3%	7.5% ± 0.3	0.4% ± 0.1	*Phymata crassipes* (100%)
	*Protaetia morio**	Coleoptera	8.3%	7.0% ± 0.4	0.1% ± 0.02	*Protaetia morio** (100%)
	unexpected parasitoid in *Forficula lesnei*	Diptera	NA	1.1% ± 0.1	0.2% ± 0.1	*Triarthria setipennis* (100%)
	chimeric reads	NA	NA	NA	3.4% ± 0.1	NA
	putative pseudogene	NA	NA	6.7%	5.9% ± 0.2	NA
	reads removed after filtering process	NA	NA	1.7%	2.0% ± 0.2	NA
Mock community MC_2	*Papilio demodocus*	Lepidoptera	14.3%	14.9% ± 0.8	27.3% ± 0.4	*Papilio demodocus* (100%)
	Thaumetopoea wilkinsoni	Lepidoptera	14.3%	19.1% ± 1.1	26.3% ± 0.5	*Thaumetopoea wilkinsoni* (100%)
	Tessaratoma papillosa	Hemiptera	14.3%	12.3% ± 0.2	17.9% ± 0.8	*Tessaratoma papillosa* (99%)
	Acorypha sp	Orthoptera	14.3%	13.3% ± 0.5	9.2% ± 0.6	unclassified (93%)
	Diabrotica virgifera	Coleoptera	14.3%	6.5% ± 0.9	6.5% ± 0.5	*Diabrotica virgifera* (100%)
	Iberorhyzobius rondensis	Coleoptera	14.3%	10.6% ± 0.6	4.3% ± 0.3	*Iberorhyzobius rondensis* (100%)
	Bactrocera dorsalis	Diptera	14.3%	9.6% ± 0.3	1.5% ± 0.2	*Bactrocera invadens/dorsalis/ philippinensis/musae/cacuminata* (100%)
	chimeric reads	NA	NA	NA	5.1% ± 0.4	NA
	putative pseudogene	NA	NA	10.5%	0.2% ± 0.03	NA
	reads removed after filtering process	NA	NA	3.2%	1.4% ± 0.4	NA

*Taxa removed for the pool after data filtering process (only 2 to 5 reads per PCR)

Table 2 Effect of the data filtering steps on the number of reads, unique sequences and (putative false and true) positive occurrences of unique sequences in the abundance table. Threshold TCC was used to filter cross-contaminations observed in the negative controls. Threshold TFA was applied to filter the false-assignments of reads to a PCR product due to the generation of mixed clusters during the sequencing. PCR triplicates were used to remove unreliable positive occurrences. NA: not applicable

			Nb of positive occurrences:			
Filter step	Nb of reads	Nb of unique sequences	for 3 to 3 PCR replicates	only for 2 to 3 PCR replicates	only for 1 to 3 PCR replicates	total
Raw data	5,835,359	5751	4969	4092	14,374	23,435
Threshold T_{CC}	5,704,150	5697	4689	3496	11,618	19,803
Threshold T_{FA}	5,684,166	5697	4466	3243	9527	17,236
PCR triplicates	5,172,708	2636	4466	NA	NA	4466

Table 3 Comparison of six DNA extraction kits used for the metabarcoding from bat faecal pellets

DNA extraction kit	Sample size	Mean number of reads per PCR	PCR failure (<500 reads)	Success of sequencing	
				Host (bat)	Prey (arthropod)**
Dneasy mericon Food Kit (Qiagen)	47	3489	7%	68%	83%
EZ-10 96 DNA Kit, Animal Samples (BioBasic)*	113	4822	14%	70%	81%
NucleoSpin 8 Plant II (Macherey-Nagel)*	95	6712	1%	84%	88%
NucleoSpin Soil (Macherey-Nagel)	9	7300	0%	67%	89%
QIAamp Fast DNA Stool Mini Kit (Qiagen)	47	2901	14%	70%	79%
ZR Fecal DNA MiniPrep (Zymo)	46	6209	1%	72%	80%

*96-well format for high throughput DNA isolation

**including unlcassified arthropod because not recorded in the sequence databases